DOCUMENT RESUME

ED 205 524 SP 018 786

AUTHOR Reed, Horace B.
TITLE The Accuracy-Meaning Tensions in Teacher Education Evaluation.
PUB DATE Jun 78
NOTE 20p.: Paper presented at the F. G. Watson Festschrift (Cambridge, MA).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Decision Making: *Educational Planning: *Evaluation Criteria: Evaluation Methods: Higher Education: Information Utilization: Program Effectiveness: *Program Evaluation: *Research Needs: *Teacher Education Programs

ABSTRACT

Teacher educators are in a state of near paralysis concerning their program evaluations. This paralysis comes from adopting wholesale a dysfunctional evaluative point of view from the dominant educational research establishment. There is a history of overestimating the current relevance of experimental behavioral science models for the gathering of accurate information and of underestimating the several meaningful criteria for evaluating teacher education. Solutions lie in relying more on alternative epistemological approaches used in history, anthropology, sociology, political science, and social psychology. Also, more weight should be given to evaluation criteria that are more inclusive than the criterion of academic pupil gains. The implications of this position require: (1) a more liberal evaluation and research curriculum for graduate students interested in teacher education; (2) a reevaluation of the teacher competence studies that are currently being winnowed by experimentally oriented reviewers; and (3) the publication and circulation of documented evaluations that illustrate more generous and useful approaches to users of teacher education information. (Author/FG)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED205524

THE ACCURACY-MEANING TENSIONS IN TEACHER EDUCATION EVALUATION

by

Horace B. Reed

(Paper presented for the F. G. Watson Festschrift, Cambridge, Mass.)

June 1978

Community Education Resource Center
School of Education
University of Massachusetts
Amherst, Massachusetts 01003

SP 018 786

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Horace B. Reed

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ABSTRACT

THE ACCURACY-MEANING TENSIONS IN TEACHER EDUCATION EVALUATION

Decisions in pre- and inservice teacher education are made daily in thousands of programs across this Nation (and the World). Many of these are made with little systematic evaluation for decision making information. Explored here is the thesis that teacher educators (among others) are in near paralysis concerning their program evaluations.

It is proposed that this paralysis comes from adopting wholesale a dysfunctional evaluation point of view from the dominant educational research establishment. There is an history of over estimating the current relevance of experimental behavioral science models for gathering usefully accurate information, and under estimating the several meaningful criteria for evaluating teacher education.

Solutions lie in broadening our beliefs as to what constitutes usefully accurate information to decision making users in teacher education. We need to rely much more on alternative epistemological approaches used in history, anthropology, sociology, political science, social psychology, perhaps literature. Also we need to give much more weight to evaluation criteria that include several different, though interacting, teacher roles that are more inclusive than the criterion of narrow academic pupil learnings.

The implications of this position would require a more liberal evaluation and research curriculum for graduate students interested in teacher education; another look at the thousands of teacher competence studies that currently have been largely winnowed by experimental science

oriented reviewers; and the publication and circulation of documented evaluations that illustrate more generous and useful approaches to users of teacher education information.

THE ACCURACY-MEANING TENSIONS IN TEACHER EDUCATION EVALUATION

INTRODUCTION

Evaluation of teacher education poses tensions between accuracy and meaning. Involved in these trade-offs are a multitude of interwoven issues, many of which are in dispute throughout evaluation in the social service fields. One consequence of the accuracy-meaning tension is to produce a near state of paralysis among teacher educators concerning program evaluation.

On the one hand teacher educators would like to have confidence that the information gathered during evaluations is trustworthy. Yet, also desired is information that is significant. Closely related to this trade-off are a multitude of other interwoven, unresolved issues. Are evaluations using the experimental science models the only or best approach? What criteria are relevant? For whom are the evaluations being made, and for what purposes? Are the teacher competence studies relevant? What are the grounds for determining accuracy? What are some alternative approaches for evaluation?

We are enmeshed in a self-defeating, socially unproductive predicament concerning evaluation of pre- and inservice teacher education programs. The paralysis stems in part from our too cautious attitudes concerning what is adequate evidence. Our respect and confidence in the direct application of experimental science epistimology to teacher education evaluation is overextended. In part, we are also boxed in by accepting socially dysfunctional, narrow criteria.

This article explores a synergy of propositions aimed at modifying our evaluation attitudes and practices. Teacher education is more a

conglomeration of conflicting theories and common sense notions, than a highly developed profession or discipline--and will remain so for the indefinite future. Teacher education is much more like social service institutions, and much less like biology or psychology or even medicine and agriculture. As a consequence we must recognize the limits as well as the potentials of experimental science's ways of knowing. We need to put more energy into utilizing and further developing additional ways of gaining information through evaluation.

Further we need to reaffirm the fact that teacher education in our culture is pluralistic in values, as well as in means. Axiological concerns must not be eclipsed by unrealistic epistemological guidelines. And concomitantly, we should keep in mind that important decisions in teacher education are being made daily, regardless of whether we have systematic evaluations to provide information for such decisions. Finally then, despite all the necessary caveats, we will do ourselves and our publics a major service if we overcome our "accuracy-meaning" paralysis. The solutions require enhancing decisions relevant to teacher education by evaluations that provide more realistic descriptions of accuracy and more meaningful kinds of information.

INSUFFICIENCY OF EVALUATIONS

It is puzzling that there have been relatively few published accounts of evaluations of pre- and inservice teacher education programs, given the thousands of programs in operation in this and other countries. These programs differ on many dimensions: in theories, goals, components, implementations, target populations, contextual situations. The ultimate

impact that program participants (pre- and inservice teachers) have on school learners is of major interest to the social order. Yet the information bank available to us indicates that relatively few programs have publicly documented in professional journals descriptions of systematic program evaluations. Annual American Educational Research Association (AERA) meetings rarely include reports of teacher education program evaluations.[1]

Conjecturing, it may be that preservice teacher education programs are protected by the mantel of higher education: other professional and liberal arts university departments also seldom undertake systematic evaluations. Inservice programs have only recently become of major interest to teacher educators, thus may not have been seen as an important area for evaluation.

More to the point may be the dysfunctional side effects produced by the dominant experimental science models for evaluation research. Staffs of teacher education programs may feel it is pointless to attempt to apply these models to their programs, given the complexities inherent in teacher education. Also operating is a prevalent belief that the only respectable criteria for teacher education effectiveness shall be pupil learning in the schools. Exclusive insistence on this criteria is a strong deterent to evaluation efforts. There are undoubtedly many other reasons for the avoidance behavior of teacher educators towards program evaluation. But these later are probably similar to the avoidance manifested by other educational groups.

LIMITS OF THE EXPERIMENTAL SCIENCE MODELS

The dysfunction of an over-insistence on experimental science models in teacher education evaluation research rests in part on an overestimate of the level of development of education as a discipline and profession. Further, it underestimates the usefulness of other evaluation models for decision makers concerned with teacher education implementation.

On the first point, experimental science research approaches require a level of sophistication well beyond the knowledge base available in teacher education. We are hardly a semi-profession, given our inadequate, conflicting theories; our meager information connecting means and ends; our controversies over objectives; our primitive measuring instruments; our inability to establish controls; our lack of information concerning contextual, intervening, and interacting variables (2).

In some instances using the experimental science models on selected problems that may have eventual relevant to teacher education evaluation is appropriate and essential. But that is a different endeavor from here-and-now evaluation of complex, ongoing teacher education programs. Also some vigorous approximations to current evaluation of teacher education, using exacting models, may be worth continued exploration in those few instances where extraordinary resources are available. But very few institutions with pre- and inservice programs have such resources.

The intention here is not to fault the underlying logic of natural science inquiry. Granted such an attack can reasonably be made (3). There are alternative assumptions, historical and current, for obtaining proof of explanations about the universe. Natural science can be examined as to its leaps of faith, and its inherent ultimate reliance on the

idiosincratic nervous systems of all those who have helped develop this system of knowing and of knowledge. But all these issues are common to every system man invents to help explain. What is intended is to question the unqualified appropriateness of certain features of the experimental sicence model to teacher education, i.e., proof of cause and effect relationships, prediction, and universal generalizability of findings. These features are crucial for purposes of knowledge building in the natural sciences. They are highly desirable, but not necessarily essential, for useful evaluation in social services areas (4).

What we are proposing, then, is to avoid the misuse of the scientific mode. We are suggesting that many of the premises, and many of the techniques, of the natural sciences can be usefully applied to evaluation in teacher education. This need not and should not require that the experimental model, or even approximations to it such as the quasi-experimental and multivariate approaches (5), be accepted as the main approaches in teacher education evaluation. What can be borrowed are such premises as integrity, replication, communication clarity, order, systematic efforts, non-causal relationships of events, empirical basis, hierarchy of abstractions, coding of information, category system.

THE ISSUE OF ACCURACY

Under what conditions is information accurate, worth trusting, reliable? Probably no information is unqualifiedly accurate for all people throughout all time. There are degrees of accuracy for any one person, that may differ from other persons, at any given time and for any piece of information. Yet in making decisions in social situations, some level of agreement is reached among relevant people concerning the

accuracy of information used in decisions. Varied bases are used by people for reaching agreement on the degree of accuracy of information they use in decision making. Throughout recorded history among all cultures there have been contending philosophical-religious positions that attempt to provide conviction as to these bases. In recent centuries, the natural sciences have increasingly held high appeal to many people throughout the world.

In our culture, some approximations to the natural science epistemological premises are generally accepted by most users. In decisions concerning legal matters, public health, legislation, social work, education, etc., many of the sciences' assumptions are accepted. But it is the users of information who have to be the judges as to what level of accuracy will be useful to them in decision making (not the producers of information, nor the thereticians backing up the producers). In teacher education evaluation literature, we have tended to place the users of evaluation information in a secondary role, letting educational research specialists call the tune as to what is acceptable evidence. Whether the user refers to others for support in making judgments about information accuracy is not the point; it is assumed such will be the case (6). The point is that there is no necessary common ground that all individual users can be guaranteed to use, or must use, in determinng their judgments.

One consequence of a user-based position is that teacher education evaluators should have some leads as to what various relevant decision makers (users) will probably want as sufficiently accurate information. This means that evaluators should recognize that there are varying degrees of information accuracy within any given system of epistemology. It may

not be the evaluators sole responsibility to decide whether the level of accuracy of information is useful or not--that may also be a role for the user.

It can be argued that imprecise information is worse than none. That all our educational research efforts should go towards knowledge building that is based on exacting experimental science models, so that our evaluations will provide highly reliable, valid information for decision making. This implies that until we have such a knowledge base in education, decisions in teacher education will be made through private idiosyncratic information held by each decision maker, with information sources being of unknown accuracy, achieved by unknown knowledge approaches. If that implication is not intended then we are forced into some workable middle ground which can be used in practical situations by specific decision makers: teacher education staffs, university administrators, school systems, state and federal education departments, teacher organizations, parent groups, and others.

RELEVANCE OF TEACHER COMPETENCE STUDIES

One extensive body of research in education has significant implications for teacher education evaluation. This is the accumulated work in the areas of teacher competence studies. There are now several thousand studies falling within this rubric, which have been carried out over the past several decades.

These studies are directly relevant to teacher education evaluation because this research includes efforts to relate teacher characteristics to pupil, classroom, and school effects. Teacher pre- and inservice programs are intended to have a bearing on teacher characteristics. Thus

if teacher competence studies shed light on desirable teacher qualities, these studies provide information useful for decision making in teacher education programs.(7).

Reviews of these studies that have received high visibility conclude that relatively little can yet be said with any confidence concerning acceptable evidence for identifying desirable teacher characteristics and behaviors. One major theme in many of these reviews is the insistence on learner gains (usually narrowly defined academic gains) as the main important criteria for teacher competence studies. And there is further insistence on the experimental science research models to obtain information. This latter requirement may be less restrictive and potentially less dysfunctional than the guidelines established for criteria.

There is no disagreement in this essay with the belief that teacher education programs should be concerned with evidence that their pre- and inservice teacher education programs do increase competencies which enhance school learnings. What is in contention is whether that is a sufficient criterion; and whether a highly restricted standard of acceptable evidence for that criteria is the only evidence that has practical utility.

One reason for researchers focusing on pupil gains as the basic criterion is that it fits the behavioral science input-output models. A consequence is to ignore the complexity of how teachers might exert influences on learners. In addition to the warranted importance put on what goes on within each individual classroom, there is also the influence of the school "atmosphere" on pupils. This school effect is in part a function of interactions of teacher-teachers, teacher-administrators, teacher-community, and all of these with pupils in and out of specific classrooms. Focusing just on individual teacher classroom effects on pupil gains generally ignores those teacher characteristics and behaviors

relevant to teachers' other roles.

Evaluation of teacher education programs should include the use of criteria that recognize the varied major roles teachers are expected to fulfill.(8) The evaluation designs appropriate for this complexity of criteria may call for gaining information about variables that do not now lend themselves to a high level of accurate numerical measuring and coding nor to resultant sophisticated statistical analysis.

Even within the domain of pupil gain criteria, the teacher competence studies have been critically reviewed using a rather narrow range of learnings as measures of gain. This may be due largely to the fact that gain scores require measures which have high validity and reliability. This narrows the field largely to basic skills. The host of additional learner objectives--complex cognitive, social, humanistic, motivational, psychological--may be less readily measured if the same accuracy restrictions are used.

It seems likely that the very large bank of information available from the decades of teacher competence studies has been too hastily winnowed. Too much may have been discarded as chaff, by the biases of many reviewers. These biases include a too rigid insistence on a behavioral science model of research and of evaluation; and on a philosophically barren set of beliefs as to valued learnings. More informed decisions makers in education have been chosen.

FREEING UP EVALUATION APPROACHES

A number of ways for gathering information can be further developed and implemented, that would be useful to decision makers concerned with teacher education. (9) An example may help illustrate. A lay school board is considering a proposal for an inservice program for their teachers. They are the primary users in this example. The program has been submitted

jointly by the administration and the local teachers organization, with the state university providing resource help. The proposal includes a plan for an evaluation of the inservice program.

A first evaluation step for the school board and the proposers is to go over the evaluation plan and see if it is clear as regards language and meaning. It may be that considerable time and effort will have to be given to this initial step. The plan submitted may include technical language which requires explanation to the board. The plan may also include ambiguities that are uncovered during this communication stage.

It is probable that during this first step, questions will arise by the school board such as whether the purposes of the inservice program are stated fully enough, or whether all the major purposes are included. Other questions may be raised concerning which components of the inservice proposal are expected to encourage selected purposes. Concern may be expressed about the sources of the techniques for data gathering. Or for how the information that is obtained will be analyzed. Issues of time, cost, disruption of schooling, etc. may be discussed.

In constructing the evaluation plan for this proposed inservice program, there are a large number of choices available concerning approaches to gathering and analyzing data. Information about some inservice program components and objectives may be most usefully gathered through means that require largely verbal descriptions, with few or no areas of information that can be meaningfully quantified. Other components and objectives of the program may readily lend themselves to quantifiable information, with less use of verbal descriptions. Respect for both forms of description opens up many sources for obtaining information, such as teacher inservice program staff, program participants, the participants' classroom students,

non-participating teachers and administrators working in the same school system, outside observers, parents of school children. Also opened up are several ways to gain information from varied sources, such as observation schedules, questionnaires, case studies, interviews, unobtrusive measures, field notes, minutes of meetings, etc.

Each of these sources, and the information-gathering approaches, requires knowledge and care if the results are to be given credence by the users of the evaluation. The degree of accuracy is not necessarily a function of whether the description of the information is qualitative or quantitative, but whether the users have confidence in the information gathering procedures.

School board members and the inservice program proposers should come to an agreement on how much credence is to be given to information gathered through the evaluation plan. This credence covers both the degree of accuracy and the meaningfulness of the information.

One possible result of this sharing process between evaluation planners and the users (the school board) may be to not go ahead with the inservice proposal. Or to go ahead with it, but to not attempt the evaluation. Even if the evaluation plan is dropped as a result of this clarification process, the positive effects of this single evaluation step on the future implementation of the inservice proposal can include heightened awareness among all parties as to intentions, expectations, problems likely to occur, questions left unanswered, conflicts that will have to be faced, etc.

SOME IMPLICATIONS

We have taken the position that the accuracy-meaning tensions can be lived with in teacher education evaluation. To do so requires a

reassessment of what are useful levels of accurate information and what are meaningful sorts of information for decision making. Useful accuracy implies an expanded range of ways to gain information, beyond the experimental science designs. Meaningful information implies an expanded range of teacher roles that are important objectives in teacher education, beyond the academic pupil learning criteria.

A great deal of evaluation in pre- and inservice teacher education must be carried out by the staffs of these programs. Only occasionally are there resources available for outside evaluators. If there is to be a marked increase in frequent, imaginative, widespread evaluations of the nations' teacher education programs, there will have to be extensive changes in emphasis in some of our university graduate programs that prepare staff for pre- and inservice programs. A negative inference is that graduate students should not be so exclusively reared on the sancity of experimental behavioral science. A positive conclusion is that graduate students should be helped to gain competence in a broad range of information gathering approaches and types of analyses used in the disciplines of history, anthropology, sociology, political science, social psychology, and psychology. It might be helpful to take seriously the usefulness and power of literary descriptions and analysis of events. And they should be helped to understand the complicated web of interacting roles of teachers, that provide the clues to the meanings of teacher education.

A further implication is that we need to document and circulate in useable form extensive, thoughtful collections of teacher education evaluations. These should provide specific examples of the very large

range of techniques, sources of information, instruments, types of criteria, modes of analysis, etc. that can be put together to provide information for differing users in decision making. These would be examples that go beyond the experimental behavioral science modes. The results or findings of any specific evaluation, using this more inclusive set of approaches, may not be highly generalizable to other users. The ways the information is gathered and analyzed, and the types of objectives studied can have extensive generalizability.

FOOTNOTES

1 For example, there is almost no discussion of teacher education-evaluation in the recent major treatise, see Elmer L. Stonening and Marcia Guttentag (Eds.), Handbook of Evaluation Research, Volumes I and II (Beverly Hills, Calif.: Sage Publications, 1975).

2 A sobering analysis of teacher education as a profession has been presented by Robert B. Howsam, et al., Educating a Profession (Washington, D.C.: American Association of Colleges for Teacher Education, 1976).

3 The critical literature on the strengths and limits of experimental science is extensive and rewarding. See these as a sample; John M. Johnson, Doing Field Research (New York, N.Y.: Free Press, 1975); Ernest Nagel, The Structure of Science: Problems in the Logic of Scientific Explanation (New York, N.Y.: Harcourt, Brace and World, Inc., 1961); Herbert C. Kelman, A Time to Speak (San Francisco, Calif.: Jossey-Bass Inc., 1968); Irwin D. J. Bross, Design for Decision (New York, N.Y.: Macmillan Comp., 1953).

4. It might be enlightening to see the results of a sociological study of the influence of educational researchers' ambiance on teacher education staff in our culture. The high status of the AERA organization among university education departments is one clue to the pervasiveness of that group's emphasis on the more rigorous modes of knowledge building. A reasonable hypothesis is that those involved in educational research and evaluation are attracted to it in part because it offers the intellectual and emotional security of high certainty concerning one's efforts. This in contrast to the ambiguity, unresolved complexity, and degree of faith in one's efforts inherent in the operations of pre- and inservice teacher education.

5 A persuasive description of those two approaches are presented by G. C. Helmstadter, Research Concepts in Human Behavior (New York, N.Y.: Appleton-Century-Crofts, 1970); William W. Cooley and Paul R. Lohnis, Evaluation Research in Education (New York, N.Y.: Irvington Publishers, Inc., 1976).

6 Users of information gained through evaluation do not make their judgments as to acceptable levels of accuracy and their decisions in a social vacuum. They are in turn accountable to their various publics; those latter have at least some vague notions of what they consider acceptable evidence. How any area of human concern is looked at by a culture (such as teacher education), with regard to credence of evidence is a highly involved historical knot, still under-explored.

7 One of the most comprehensive overviews of this, and related, area of research is covered in depth and with accompanying extensive bibliographies, in Robert M. W. Travers (Ed.), Second Handbook of Research on Teaching (Chicago, Ill.: Rand McNally and Comp., 1973). The First Handbook edited by N. L. Gage in 1963, provides further coverage that is not replicated in the 1973 project.

8 A more comprehensive set of teacher competence criteria should include not only direct effects on academic learning, but the teacher roles of guidance, involvement in school studies and innovations, school-community relations, and members of a profession.

9 A number of good leads can be sampled in the following: Herbert J. Walberg, Evaluating Educational Performance (Berkeley, Calif.: McCutchan Publishing Corp., 1974); Michael W. Apple, Michael J. Subkoviak, and Henry A. Lufter, (Eds.), Educational Evaluation: Analysis and Responsibility (Berkeley, Calif.: McCutchan Publishing Corp., 1974);

Carol H. Weiss, (Ed.), Evaluating Action Programs: Readings in Social Action and Education (Boston, Ma.: Allyn and Bacon, Inc., 1972); Ernest R. House (Ed.), School Evaluation: The Politics and Process (Berkeley, Calif.: McCutchan Publishing Corp., 1973); Herbert C. Kelman, op. cit.; L. J. Cronbach and P. Suppes (Eds.), Research for Tomorrow's Schools: Disciplined Inquiry for Education (New York, N.Y.: Macmillan, 1969).